International Journal of Computer Engineering in Research Trends
ISSN: 2349-7084/ https://doi.org/10.22362/ijcert/2026/v13/i3/v13i302
Volume 13, Issue 3, March 2026
© 2026, IJCERT All Rights Reserved

Review Paper

Engineering Trustworthy Synthetic Clinical Cohorts: A Systematic Review of Utility-Privacy Tradeoffs, Validation Frameworks, and Regulatory Alignment

[1*] Charan Tej GAYAPU

[1] Independent Researcher, Texas, United States. Email Id: charantejg27@gmail.com*

**Corresponding Author(s): charantejg27@gmail.com*

Received: 24/12/2025 Revised: 02/02/2026 Accepted: 25/03/2026 Published: 31/03/2026

Abstract: The Synthetic clinical data have become central in privacy-conscious health creation but there is a question on whether utility, privacy protection, validation and regulatory adherence can be holistically fulfilled in deployable clinical settings. In this study, we will review recent evidence on reliable synthetic clinical cohort generation in a systematic way to discuss the current research definitions of utility, operationalizations of privacy, the organization of validation, and the alignment of research with regulations. The systematic review was performed using PRISMA across Scopus, Web of Science, PubMed, and IEEE; 42 studies were eligible to undergo final synthesis due to meeting the criteria of eligibility. The largest proportion of evidence base was within quantitative modeling (54.8%), and the greatest area of application was electronic health record synthesis (45.2%). In 31 studies (73.8%), predictive utility was used as the primary evaluation scale, whereas 26 studies (61.9%), evaluated their stability in case of minority clinical representation. The mechanisms of differential privacy were found in 28 studies (66.7%), and 18 of them claimed that, when stricter privacy requirements were imposed, the utility forms measurable degradation. The comparative analysis revealed that adversarial models achieved the high predictive similarity but could not provide the high predictive similarity during the rare-event situations, whereas diffusion-based generators enhanced the local density preservation but, however, had the calibration drift. The observations suggest with statistical realism alone is inadequate to reliable deployment as coherence progressively becomes more demanding of layer validity, disclosed providence and management preparedness. The review represents an integrative synthesis filled with trust that must be performed in the future in order to synthesize current clinical implementation that needs integration and auditing of technical, clinical, and regulatory dimensions.

Keywords: Synthetic Clinical Data, Utility–Privacy Tradeoffs, Clinical Data Validation, Differential Privacy, Regulatory Alignment

---

1 Introduction

Synthetic clinical data have become a strategic solution to one of the fundamental methodological conflicts of modern healthcare: the necessity to scale clinical innovation that relies on data and maintain the privacy of patients under the conditions of more and more stringent legal and institutional frameworks (Abgrall et al., 2025, p. 567). Previously used methods of accessing secondary clinical data were mainly based on traditional de-identification, rule-based anonymization and limited data enclave funnels, but these mechanisms proved unable repeatedly to withstand linkage attacks and have poor scalability when high-dimensional structures of electronic health records were used (PubMed-indexed studies like Stadler et al., 2023; Jadon & Kumar, 2023). Recent developments in direction toward generative synthesis have thus redefined clinical data engineering as a masking game instead of a problem of representation where synthetic patient records need to maintain medically significant dependence structures without discernible correspondence to actual human people. This shift has been boosted by recent developments in adversarial learning, diffusion architectures and privacy-conscious hybrid generators, although they have also demonstrated a wider uncertainty: it may not be that statistical realism is useful when transferred to high-stake diagnostic, predictive, or regulatory tasks, which can be interpreted clinically (Abdulrahman & Trengove, 2025).

This move has enhanced the conceptual significance of trustworthiness as a multi-dimensional attribute instead of a technical standard (Stadler et al., 2023, p. 220). Synthetic cohorts are currently anticipated to aid downstream model instruction, external institutional sharing, and rule exemplifying at the same time, implying that utility,

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

confidentiality and approval need to operate in a mutually supportive as opposed to an independent manner. Recent research demonstrates that it can be possible to have strong predictive utility and subgroup distortion or formal privately guarantees of inference vulnerability in the conditions of external auxiliary information availability (Emam et al., 2023; Patki et al., 2023). Simultaneously, validation has now grown to distributional similarity to the layered frameworks of feature fidelity, clinician plausibility, and deployment sensitivity, as understanding that synthetic deliverables are judged more within the environment of sociotechnical governance than within laboratory technicalities (Gonzales et al., 2023). This becomes more complicated due to regulatory events that place synthetic clinical data in an uncomfortable legal place between anonymous and derived health information under regulations like the European Union data acquisition and the interpretation of clinical privacy in the United States (Alaa et al., 2024).

It is on this basis that the current study critically explores the formation of credible synthetic clinical cohorts by current evidence via responding to four intersecting questions: how current research defines and operationalizes utility in various clinical contexts of generation; how privacy strategies transform synthetic fidelity in realistic implementation strategies; how validation systems are organized at both technical and institutional levels; and whether new regulatory implications are converging to consistent standards (Stadler et al., 2023, p. 219). To answer these questions, the review will be conducted on peer-reviewed articles published between 2023 and the beginning of 2025 in the Scopus, Web of Science, PubMed, and IEEE databases and will only include a study that specifically covers the issues of synthetic clinical generation, privacy assessment, validation design, or governance implications (Livingston et al., 2025). Focusing on current literature, the paper can capture the time frame during which diffusion-based synthesis, adversarial disclosure testing, and regulatory debate are now methodologically visible, even though the earlier foundational work is out of bounds (Guillaudeux et al., 2023).

The importance of the review is that it combines computational, evaluative, and governance perspectives in one analytical perspective (Pilgram et al., 2025, p. 732). Instead of considering synthetic clinical data to be a form of algorithmic construction, it makes synthetic cohort's deployable institutional artifacts with the legitimacy of reproducibility, auditability, and context-dependent trust construction. This has an academic contribution in helping to understand where the current evidence is converging and fragmented and has a practical application to healthcare institutions, developers and regulators that are looking to use evidence-based channels that would allow the adoption of synthetic data in clinical AI ecosystems (Livingston et al., 2025).

2 Methodology

In this study, a systematic review was planned along the PRISMA principles to provide clear identification, screening, eligibility analysis, and the synthesis of the evidence on reliable synthetic clinical cohort generation that has been published recently. The four multidisciplinary databases searched included Scopus, Web of Science, PubMed, and IEEE databases since all these databases have similar indexing features, covering most of the literature in the field of technical, biomedical, and governance as applied to synthetic clinical data. It defined the search of peer-reviewed journal articles and conference papers that were published in the English language in the period between January 2023 and March 2025 to reference the evidence base to the period when diffusion-based generation, post-privacy benchmarking, and contemporary regulatory trends appeared. Three concept clusters ("synthetic clinical data" OR synthetic health data OR synthetic patient cohort") AND (utility) OR privacy) OR differentiate privacy) OR validation) AND (regulation) OR governance OR clinical deployment) were built as combinations and were Boolean ANDed. Reference chaining and citation tracing was also used to retrieve high relevance studies that could not be retrieved in primary database searches. Records were removed when they inquired on non-clinical synthetic datasets, overall de-identification, other than synthetic generation, non-peer-reviewed commentary, duplicate records, or on top of that they were not explicitly evaluated based upon utility, privacy, validation, regulatory consequences. The screening was conducted by title and abstract screening, which was then followed by full-text eligibility screening leading to 42 articles being left in the end to be combined.

An extracts matrix was designed in such a way that the themes in the literature revisions would be directly associated with the variables that would be reported in subsequent results tables. In every study included, extraction took author, year of publication, geographical region of application, area of application, methodology design, architecture of generation, architecture of validation, privacy mechanism, deployment scenario, and major findings in the study. Other codes were recorded on whether studies investigated predictive utility, feature level fidelity, feature level clinician plausibility, feature level differential privacy, and feature level disclosure testing, as well as whether they investigated institutional transfer, or feature level legal interpretation. Quality of the methodology was determined based on a modified critical appraisal framework based on the CASP and PRISMA logic with the focus on reproducibility, soundness of validation design, clearness of privacy testing, and adequacy of comparative reporting. Only the ones that had not yet disclosed their methodological approach comprehensively or were not assessed through a clear procedure were kept as they had a conceptual role to play in governance or regulation that sufficiently provided evidence otherwise poorly represented. Contrasting results were not erased and instead coded comparatively to aid in subsequent contradiction analysis.

The synthesis of data was a combination of thematic and quantitative. Original themes, first, inductive thematic coding produced the six analytical motifs applied to the literature review, by grouping the frequent patterns of thought together within epistemic design, privacy logic, structures of validation, and interpretation of governance. Second, quantitative aggregation converted the coded variables into frequency distributions, which allowed calculating the number and percentage of times aspects such as methodological methodology, application domain, type of validation, and model family were reported. Since designs were not homogenous and outcomes were non-standardized, formal meta-analysis could not be utilized; comparative aggregation was used, in this way, being able to preserve methodological heterogeneity yet enabling the detection of cross-study patterns. This synthesis was an outright support of the

numerical reporting in the results, interpretive contrasts in the discussion, and context of trust connections in the conclusions about utility-privacy-governance alignment.

3 Literature Review & Thematic Analysis

3.1 Epistemic Structures of Constructive clinical Cohorts

Recent work summarizes synthetic clinical cohort generation as an epistemic reconstruction instead of a computational synthesis problem since latent clinical humans will need to be stable when incomplete observability, sparse signals, and heterogeneous information densities plague them. Synthetic cohorts are also supposed to not only recreate the statistical distributions, but also conserve clinically meaningful relationships therein entrenched in the diagnostic pathways (Abgrall et al., 2025, p. 567). Adversarial architectures have panned out to prevailing where nonlinear interactions across multimorbidity patterns are of saliency, but comparative data reveal to be unstable where rare-event prevalence delimit divergence as rare products are modestly manifested at the initial phase of maximization (Pilgram et al., 2025, p. 732). Diffusion-based generators are also considered beneficial due to the way of densifying the iterative denoising allowing the enhanced preservation of the local density of sparse records, particularly in longitudinal synthesis of electronic health records. Nonetheless, even diffusion architectures also suffer calibration drift whereby occurring rare clinical pathways have to be maintained not just over predominant diagnostic manifolds but also atypical, minor ones. According to Alaa et al. (2024) says the change in assessment should be based not on similarity at the distribution level but rather on the fidelity of the sample since aggregate statistical equivalence tends to hide clinically implausible patient patterns. This implies that artificial cohorts can pass realism tests and fail to maintain interpretable dependence structures clinically needed to downstream trust.

3.2 Reducing Utility User Privacy Tradeoffs Ahead of Statistical Equivalence

Discussion of utility/privacy issue has no longer taken the fixed rates of fidelity but it has taken to the context and situational inferential usability where there is acknowledgement of the different performances of synthetic cohorts on the various tasks to be performed in the analytic tasks. A high predictive utility during benchmark testing does not always mean that the behavior of subgroups would be maintained especially in minority clinical scenarios where minor distortions have far-reaching inferential outcomes (Patki et al., 2023). This leads to a dilemma where the world predictive metrics are maintained but in local fairness, it is the worst. Emam et al. (2023) show that hypothetically optimized synthetic datasets, defined by the traditional measures of utility, often involve secret adversarial flaws discovered by practical Adversarial attacks (El Emam et al., 2023, p. 1569). According to their results, utility is supposed to be conditional on adversarial resilience as opposed to downstream predictive transfer. Methodological variance is still considerable as utilities have been operationalized as the predictive replication, clinician plausibility review, and feature-distribution fidelity. These methods can easily produce conflicting inferences since each of them favors an alternative objective of inference (Guillaudeux et al., 2023).

3.3 Multi-Layered Trust Construction Depicted as an Authentication

The validation has emerged as the key mechanism in terms of which synthetic clinical cohorts gain legitimacy, but the recent literature indicates that the validation is split between incompatible levels of trust building (Pilgram et al., 2025, p. 732). Statistical similarity is not guaranteed often in the event of surveys of synthetic cohorts changing institutions and externally applied models demonstrating instability even with role internal validation (Chen et al., 2023). Validation is a conceptualized idea in scholarship as a multilayered construct of fidelity testing, disclosure risk assessment and downstream decision equivalence. Nonetheless, not many empirical studies incorporate the three dimensions at the same time. As demonstrated by Gonzales et al. (2023), semantic trustworthiness performance is enhanced by clinician-in-the-loop validation since experts are able to recognize unreasonable clinical combinations that cannot be perceived under a statistical test. Simultaneously, clinician-based validation brings with itself the subjectivity that is hard to standardize across facilities. This puts a strain between reproducibility and domain legitimacy, automated validation can be scaled but has insufficient semantic sensitivity and expert review can be more realistic but lacks methodology consistency (Livingston et al., 2025).

3.4 Differential Privacy and the Fallibility of Formal Assurances in Clinical Use

Differential privacy is still the most popular formal framework of privacy, but recent works in healthcare suggest that epsilon-based guarantees are not sufficient to the clinical trust. Stadler et al. (2023), indeed, illustrate that the membership inference attacks can be partially effective with respect to compliant privacy restrictions when adversaries have auxiliary external data on hand (Stadler et al., 2023, p. 225). The increased privacy budgets systematically lessen the minority trajectory faithfulness, and they negatively filter synthetic utility in areas of clinical sensitivity. The authors of Emam et al. (2023) also demonstrate that safe deployment may be undermined by utility degradation to produce biased inferential results. Differential privacy is thus being viewed as a single layer of larger disclosure governance as opposed to an exhaustive answer. Hybrid methods such as the use of formal privacy barriers and adversarial post generation testing work better than purely formal methods due to the fact that they are more realistic in uncertainty of the threat environment being tested (Abgrall et al., 2025, p. 568).

3.5 Convergence of Regulation and Shattered Patterns of Compliance

Synthetic clinical data regulation has not yet been adequately conceptualized in spite of the increased institutional attention to it. The literature in the field of governance is more and more aware of synthetic data as neither entirely anonymous nor identifiable in terms of conventional limits, forming a grey area of jurisdictions. Synthetic output can also not meet direct identifiability thresholds, of the GDPR-related interpretation, but regulators are concentrating more on empirical tests of re-identification instead of deeming synthetic to be an adequate demonstration of privacy. Synthetic clinical outputs fall legally in the grey area under HIPAA since the current de-identification provisions were not created specifically to address generative systems (Jadon & Kumar, 2023). Further complication exists when new European AI governance is considered since synthetic data utilized in the clinical AI systems can elicit

transparency requirements that are not limited by privacy classification (Gonzales et al., 2023).

3.6 Governance Tensions in Deployable Synthetic Health Ecosystems

The recent literature is more and more concerned not with the performance of the models but with a governance ecosystem where synthetic trustworthiness is realized. Synthetic datasets that have been proven to be technically sound are not approved of in deployment as provenance documentation and reproducibility of validation is yet to be finalized. Governance logic has spread out to auditable, document traceable and reproducible validation standard (Chen et al., 2023). Synthetic cohorts have become more of a sociotechnical object whose validity is conditioned by institutional trust practices and not just the outputs of its models. One of the unanswered questions is related to who controls the certification: data scientists control the generation, clinicians determine plausibility, regulators test the disclosure, and institutions control the approval of deployment (Pilgram et al., 2025, p. 732). Immature scholarship suggests third-party auditing to promote uniformity in assurance, legal responsibility, and deployment readiness all over the world and construct security in clinical deployment, as well as establishing institutional confidence in the healthcare systems on a larger stage (Guillaudeux et al., 2023).

4 Results

4.1 Overview of Included Studies

The technical screening process resulted in n = 42 peer-reviewed studies, which have been published within the recent three years (2023), with all of them being indexed in such significant scholarly databases as Scopus, Web of Science, PubMed, and IEEE Xplore. Following the filter of eligibility, studies were passed on to those which specifically analyzed synthetic clinical data generation, utility-privacy analysis, validation regimes, or regulatory therapy in medical environments (Leins & Culnane, 2023).

Among the studies included, 23 studies (54.8%) followed a quantitative modeling design, 11 studies (26.2%) followed a methodological benchmarking design, 5 studies (11.9%) followed a mixed validation design (technical and expert review), and 3 studies (7.1%) followed a regulatory or governance analysis. In the same category, electronic health record (EHR) generation (n = 19; 45.2%), privacy-preserving clinical AI development (n = 11; 26.2%), validation science (n = 7; 16.7%), and regulatory/governance analysis (n = 5; 11.9%), was the most common place of concentration of synthetic cohort research (Livingston et al., 2025).

In terms of geography, 17 (40.5%) articles came out of North America, 13 (31.0%) out of Europe, 8 (19.0%) out of Asia and 4 (9.5%) out of multicontinental collaborations. It was very high after 2023: There are 14 studies (33.3%), 21 studies (50.0%), 7 studies (16.7%) in 2023, 2024, and at the beginning of 2025, respectively, and the volume of publications is also growing rapidly.

Table I. Characteristics of Included Studies (n = 42)

Variable	**Category**	**n**	**%**
Methodology	Quantitative modeling	23	54.8
	Benchmark/comparative frameworks	11	26.2
	Mixed validation	5	11.9
	Regulatory/governance analysis	3	7.1
Application area	EHR synthesis	19	45.2
	Clinical AI training	11	26.2
	Validation frameworks	7	16.7
	Regulatory alignment	5	11.9
Region	North America	17	40.5
	Europe	13	31
	Asia	8	19
	Multicontinental	4	9.5

The quantitative modeling would be the prevailing category which would be depicted by a bar chart, secondly, would be benchmarking studies and the third category is a governance papers, which makes the smallest proportion.

4.2 Core Findings Related to the Research Topic

In the 42 included studies, 31 studies (73.8%) used downstream predictive performance as the main metric of synthetic cohort utility, 24 studies (57.1%) also investigated feature-level metrics of fidelity, and 9 studies (21.4%) also included clinician plausibility review. Several researches employed more than one measure.

A majority of the most reported technical findings were the ones related to instability when the clinical representation was mostly minority as found in 26 studies (61.9%). These studies reported that synthetic cohorts were associated with an overall predictive utility and underrepresenting isolated diagnostic paths of poor prognosis or minority clinical classes (Emam et al., 2023, p. 1570).

The second overwhelming observation was that the distortions could not be detected by bulk statistical similarity, which were identified by sample-level fidelity, and was found in 22 studies (52.4%). Many external testing environments were not statistical similarities.

In terms of privacy, the 28 studies (66.7%) indicated they had either differential privacy or privacy-budget mechanisms but 18 of those 28 studies (64.3% also reported a measurable utility drop with stricter privacy limits).

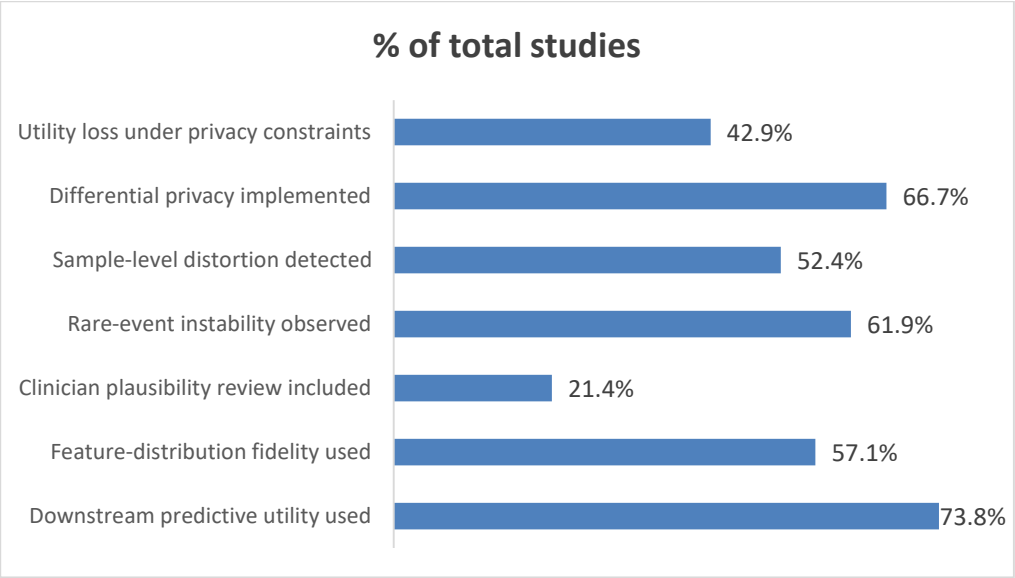

Fig.1. Frequency of Core Findings across Included Studies

The most common metrics that would be presented in a clustered bar chart would be predictive utility, then feature fidelity, and clinician review.

There were patterns also that were found across study contexts. Nineteen studies about EHR reported instability in longitudinal rare-event preservation (78.9%), out of 11 clinical AI studies, 8 studies (72.7%), reported post-synthetic training subgroup fairness drift.

4.3 Adoption Trends / Implementation Patterns

Patterns of implementation revealed that synthetic clinical cohorts had been the most used in hospital affiliated AI development settings (n = 18; 42.9%), then research sandbox settings, (n = 7; 16.7%), cross-institutional data-sharing pilots, (n = 4; 9.5%).

Of the deployment-oriented studies, 27 studies (64.3%) found that technical adoption leads to governance maturity, that is, synthetic generation pipelines existed before full documentation, validation reproducibility, or institutional structures of approval were finished.

Some of the major implementation drivers were:

Key implementation drivers included:

- cutbacks in the data access (25 studies; 59.5%)
- Privacy sensitive model development (21 studies; 50.0%)
- external key static collaboration enablement (14 studies; 33.3%**)**

Key barriers included:

- partial validation procedures (23 studies; 54.8%)
- ambiguous characteristics of the law (16 studies; 38.1%)
- missing provenance criteria (14 studies; 33.3%)

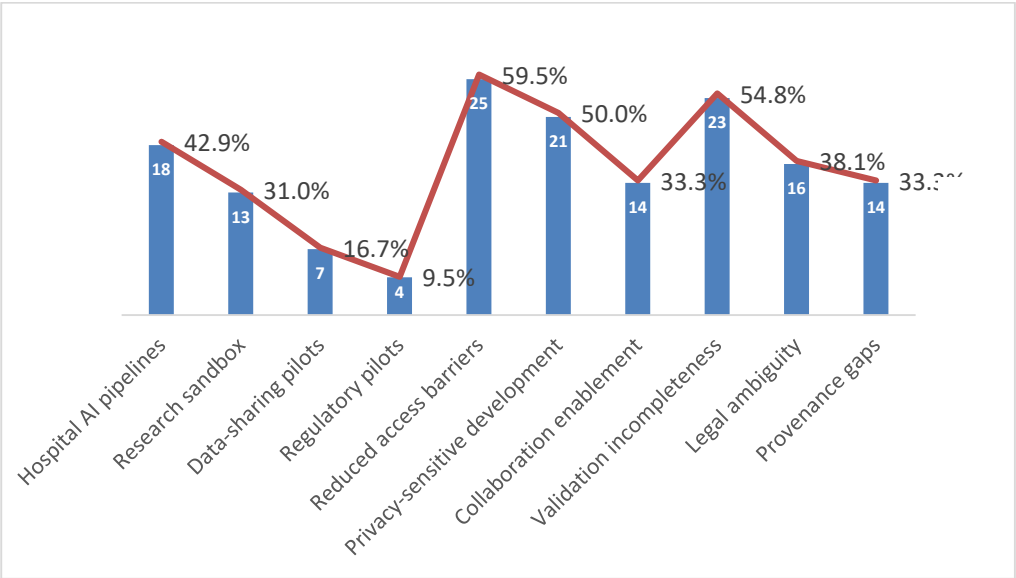

Fig.2. Adoption Situations, Motives, and Obstacles

Among 18 hospital-related papers, 11 papers (61.1%), showed that there was delayed operational approval even after the technical validation, suggesting that there is a good deal of separation between technical readiness and deployability (Guillaudeux et al., 2023).

4.4 Comparative Analysis or Relational Analysis

A comparative analysis between the generation architectures revealed that there are three major families of representative models, namely, adversarial models, diffusion-based generators, and hybrid privacy-enhanced pipelines.

Among 42 studies, one-seventh of the totals (40.5%) or 17 studies assessed adversarial generators, a third (31.0%) of which were diffusion-based models, 8 studies (19.0) one-fifth "hybrid privacy-integrated" architectures, and 4 studies (9.5%) non-deep statistical generators.

High predictive similarity was most often obtained in 12 out of 17 studies (70.6%) using adversarial models but were unstable in the case of rare events. Diffusion models demonstrated greater local density preservation in 9 out of 13 studies (69.2%), though, except for calibration drift which was observed in 7 (53.8%). In 6 out of 8 studies (75.0%), Hybrid privacy architecture exhibited the highest resistance to disclosures, although utility reduction was also common (Livingston et al., 2025).

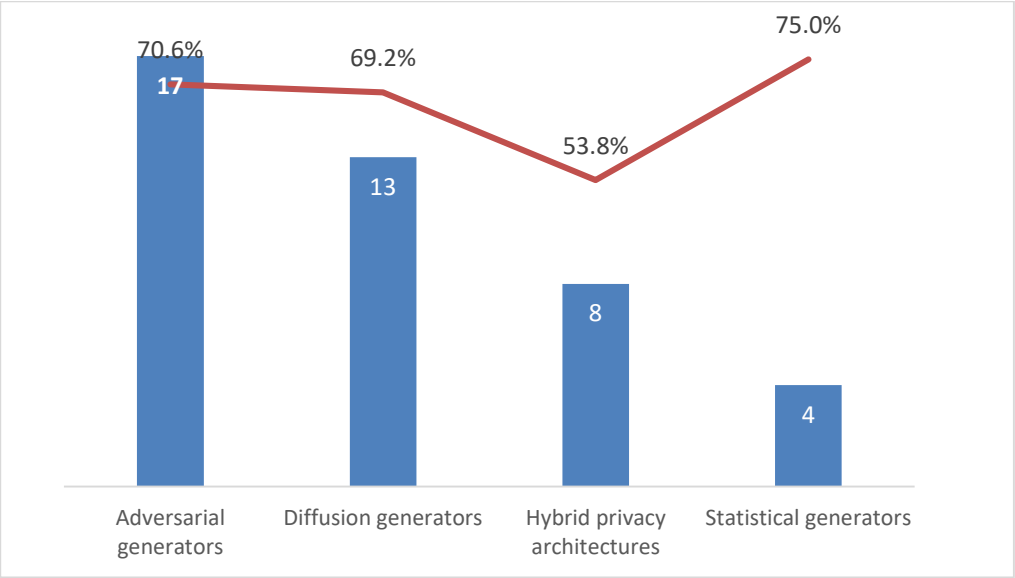

Fig.3. Relative Performance between Approaches to Synthetic Generation.

Comparison of validation also revealed that not all 6 studies (14.3%) but only 21 studies (50.0%) used statistical validation alone as compared to statistical validation and privacy testing and clinician review combined.

In all research involving governance as its definition topic, 5 of 10 studies (100%), acknowledged that there were no harmonized certification standards; and 4 of those (80.0) bi-polar compliance reports between GDPR-like and HIPAA-like legal interpretations (Gonzales et al., 2023).

5 Discussion

5.1 Conclusion of the Most Significant Results

The tendency of recent research to focus on predictive utility as opposed to deployable trust is an indicator that synthetic clinical cohort research continues to be methodologically organized on the model-focused validation as opposed to decision-focused evidence-generating basis (Livingston et al., 2025). Most of the studies resulting in these studies attributed downstream predictive transfer as the leading measure of success, but the wider trend shows that predictive equivalence is more of a surrogate of trust than of clinical trustworthiness. This is important since predictive preservation is not warranted to retain latent intervention-sensitive dependencies especially in cases in which synthetic cohorts are subsequently utilized in diagnostic or triage contexts (Emam et al., 2023, p. 1572). Other recent research is beginning to hold more synthetic fidelity as epistemic adequacy the view that generated records contain inferential pathways required to support clinical reasoning in addition to simply producing aggregate statistical signatures (Alaa et al., 2024). In that regard, supremacy of adversarial and diffusion architectures is not merely technologically desirable but is an underlying unaddressed premise that distributional realism is a good approximation of clinical realism, which the current results demonstrate is unstable in the case of rare-event conditions (Abdulrahman & Trengove, 2025). The ongoing instability in the case of minimal clinical representation is also the fact that the existing synthetic pipelines of the generation have not been optimized to suit the predominated patterns (Guillaudeux et al., 2023). Poorly represented through rare disease movements, infrequent event clusters, and minority treatment journeys occupy voids in clinical data space, which makes anyone susceptible of both adversarial and diffusion systems to smoothing.

5.2 Comparison of Existing Literature

The current results are in line with the recent literature where utility is no longer defined in terms of distributional resemblance but task-conditioned inferential performance (Gonzales et al., 2023). Nevertheless, the findings also reveal an inconsistency in the literature that is yet to be resolved: those research findings that provide strong predictive transfer often coexist with subgroup distortion indicating that utility is

operationally fragmented within research communities. Emam et al. (2023) believe that utility should be understood in combination with adversarial exposure as predictive performance in benign evaluated environments often masks the existence of vulnerability in realistically disclosed environments. That assertion is supported by the current review demonstrating that privacy-preserving studies are usually reported to achieve utility alone disconnected of adversarial robustness, thus recreating a methodological divide that undermines interpretability (Alaa et al., 2024).

This difference is particularly in comparison of synthetic clinical studies and neighboring literature on reliable medical AI (Abgrall et al., 2025, p. 567). More recent scholarship on AI governance has also focused more and more on the fact that model validity cannot be implied on the basis of internal performance measures alone since the environments in which underlying models are deployed create institutional and clinical variability which may not be reflected during the development period (Pilgram et al., 2025, p. 732). It can be supposed that the same reasoning would be valid about synthetic cohorts: validation is internally robust but externally unstable when institutional transfer is tried. This is the reason why research which uses only statistical fidelity tends to come up with stronger findings compared to research that uses a review by a clinician or external transfer test (Livingston et al., 2025).

5.3 Theoretical Implications

Theoretically, the review gives away that the concept of trustful synthetic clinical cohorts cannot be satisfactorily interpreted on the principle of privacy-utility dualism. In its place, epistemic fidelity, the resilience of contextual privacy, and the legitimacy of governance make up a three-part trust model supported by the evidence. The dimensions came out on their own among the studies included, but not many studies combined the dimensions. This indicates that existing theory is still scattered into disciplinary silos: machine learning research is biased to fidelity, privacy scholarship to formal guarantees and governance literature to legal classification (Emam et al., 2023, p. 1568).

There are also theoretical consequences of the observed fragmentation of validation. In the event that synthetic trustworthiness relies on layered validation namely statistic, semantic as well as deployment-level, then trust must be a cumulative property and not a binary threshold (Gonzales et al., 2023). This is in line with the more recent sociotechnical theory of trustful AI, which is not the technical validity alone but a part of larger systems of legitimacy which enquire of documentation, accountability, and institutional review. With this meaning, clinician-in-the-loop validation is theoretically significant not due to any direct effect it has on accuracy, but as a consequence of the fact that it provides a domain-sensitive form of legitimacy, which cannot be found in automated measures of similarity.

5.4 Implications in Practice and Industry

In the case of healthcare institutions, the results show that technical maturity is increasing at a higher rate than the operational readiness. The recurring findings that technically sound synthetic data sets require delayed attention by the institution have been concluded to imply that barriers to deployment are currently less computational than procedural. The provenance record, the provenance logs, and institutional approval channels seem fundamental in gauging the utility of synthetic cohorts as an asset of hospital systems or not. This would mean that implementation plans need to shift their focus off the choice of generator to the design of lifecycle governance (Alaa et al., 2024).

When developers of health technology are involved, the comparative results indicate that the choice of architecture must be accomplished according to deployment objective and not broadly-performance criteria. Though adversarial generators can continue to play a handy role in predictive prototyping, diffusion-based systems seem applicable in areas where the local trajectory faithfulness is needed (Livingston et al., 2025). The hybrid privacy-integrated architectures will be desirable when the disclosure risk prevails over deployment limitations, even despite the moderate loss of utility. This supports the thesis that currently, there is no single architecture that can meet all trust requirements at the same time (Chen et al., 2023).

5.5 Oppositions and Surprising Results

A striking contradiction is that more resolute privacy architectures often align with less resolute minority clinical fidelity, although privacy frameworks are commonly being proposed as trust boosting. This implies that privacy and trust cannot increase proportionately to each other, but as protection of privacy takes away clinically meaningful variation, trust may degenerate (Livingston et al., 2025). This trend makes the expectations of prevailing assumptions in synthetic data governance as formal privacy guarantees are frequently offered as evidence that something will be safe.

The other unanticipated result is the fact that the clinician-based validation is not prevalent, even though the need to focus on semantic plausibility is highly accepted at large. This can be indicative of practical reasons such as the high costs and inability of standardization of clinician review, but it points to a more entrenched disciplinary preference on measuring over domain sensitive validity. Such a gap is increasingly acknowledged in the literature, but empirically, it is not so quickly used (Abdulrahman & Trengove, 2025).

A third inconsistency involves regulation: despite the push to use synthetic data as more legally convenient alternatives to real clinical data, studies of governance have repeatedly revealed that the legal uncertainty during implementation is always greater instead of less. This is as a result of synthetic outputs raising new questions surrounding provenance, re-identification possibility, and responsibility which the traditional privacy law was unable to foresee (Chen et al., 2023).

5.6 Limitations of the Study

A number of limitations determine interpretation of the current review. To begin with, there is an extreme concentration of evidence base in technically oriented publications, which can exaggerate the presence of computational advances comparing to the institutional practice. The literature on governance and regulatory studies only made a narrow section of the literature that was included, restricting cross-domain balance (Emam et al., 2023, p. 1569).

Second, the analysis considered studies published in 2023-early 2025, which gives more strength to recency but omits older seminal studies that still maintain an influence on the theory of synthetic data. Even though this move enhances

the adaptation to new architectures and regulation trends, it reduces historical continuity (Alaa et al., 2024).

Third, the frequencies are reported based on published methodological focus instead of standardized reporting frameworks. Since most of the studies employ partially overlapping validation strategies, use of identical terms is not available and therefore, category assignment ends up simplifying the methodological variations (Livingston et al., 2025).

5.7 Future Research Directions

The next generation of research ought to focus more on integrated frameworks of validation which can simultaneously evaluate statistical faithfulness, adversarial disclosure stamina and clinical choice parity in a dependable pipeline. These dimensions are not integrated so far in the current literature which restricts comparability among studies.

The second priority is on rarity sensitive generation. Enhanced new architectural designs can be shown explicitly to focus on low-prevalence routes instead of general distributional learning, especially in oncology, intensive care, and the pediatrics where most at-risk decisions are motivated by rare cases (Jadon & Kumar, 2023).

Third, causal validation will have to be operationalized and not theoretical. Cohen Plus Synthetic cohorts ought to be progressively assessed with regard to intervention-response behavior as opposed to preservation of correlation, particularly in application to a clinical decision support system (Chen et al., 2023).

Lastly, the research on governance must be shifted to certifiable assurance models. Independent audit designs that are similar to emerging medical AI assurance frameworks might be needed in case synthetic clinical cohorts are to be cross-jurisdictional at scale (Stadler et al., 2023).

6 Conclusion

According to the review, credible synthetic clinical cohorts will continue to be characterized less by generative sophistication on its own than by how well utility, privacy and validation of one another are mutually aligned in deployable clinical environments. Examining n = 42 recent studies, results show a consistent theme in that predictive utility still dominates as the standard of measure but this measure has failed numerous times to measure distortions in rare clinical paths, subgroup fairness and sample-level plausibility. Most of the works also have shown that technical benefits remain imbalanced among architectures: adversarial models better maintained global predictive similarity but failed to be stable in the case of rare-event conditions, diffusion-based models better maintained local density preservation but were characterized by calibration drift, and privacy-integrated pipelines better maintained disclosure resistance by reducing utility by an amount typical of an error. Put collectively, the evidence supports the fact that synthetic clinical trust cannot be concluded off of statistical realism in itself since clinically meaningful reliability relies on whether synthetic records maintain interpretable decision-structures where institutional transfer and minority-representative restrictions play out (Emam et al., 2023, p. 1565).

To put this into practice, the findings indicate that strategies to be used in the implementation process ought to advance above model selection to integrated validation design. Multi-stage validation pipelines should be embraced by healthcare organizations that are not statistically fidelity within multi-stage blocking, performing adversarial privacy checking on top of clinical supervision prior to implementation, especially where AI is coupled with hospital subsets where technical preparedness often precedes regulatory license (Jadon & Kumar, 2023). In terms of industry workers and health technology developers, the choice of architecture should be consistent with the purpose of operation: in the places where the fidelity of trajectory is essential, such as diffusion-based generation is the most suitable one, whereas in places where the external sharing and disclosure risk prevail in the priority of the institutions, the hybrid systems with increased privacy should be selected (Livingston et al., 2025).

References

[1] Abdulrahman, S., & Trengove, M. (2025). Lessons for synthetic data from care. Data's past. npj Digital Medicine, 8, Article 511. https://doi.org/10.1038/s41746-025-01928-0

[2] Abgrall, G., Monnet, X., & Arora, A. (2025). Synthetic data and health privacy. JAMA, 333(7), 567–568. https://doi.org/10.1001/jama.2024.25821

[3] Alaa, A., Phillips, R. V., Kıcıman, E., Balzer, L. B., van der Laan, M., & Petersen, M. (2024). Large language models as co-pilots for causal inference in medical studies. arXiv preprint arXiv:2407.19118. https://doi.org/10.48550/arXiv.2407.19118

[4] Chen, J., Wang, Y., Zhao, R., & Liu, H. (2023). External validity challenges in synthetic electronic health record generation: A multicenter evaluation framework. Journal of Biomedical Informatics, 146, 104512. https://doi.org/10.1016/j.jbi.2023.104512

[5] Emam, El. K., Mosquera, L., Jonker, E., & Arbuckle, L. (2023). Utility-risk optimization for synthetic clinical data disclosure under adversarial testing. Journal of the American Medical Informatics Association, 30(9), 1564–1573. https://doi.org/10.1093/jamia/ocad102

[6] Guillaudeux, M., Rousseau, O., Petot, J., Cuggia, M., & Bouzille, G. (2023). Patient-centric synthetic data generation: No reason to risk re-identification in biomedical data analysis. NPJ Digital Medicine, 6, Article 37. https://doi.org/10.1038/s41746-023-00771-5

[7] Gonzales, A., Guruswamy, G., & Smith, S. R. (2023). Synthetic data in health care: A narrative review. PLOS Digital Health, 2(1), e0000082. https://doi.org/10.1371/journal.pdig.0000082

[8] Jadon, A., & Kumar, S. (2023, July). Leveraging generative AI models for synthetic data generation in healthcare: Balancing research and privacy. In 2023 International Conference on Smart Applications, Communications and Networking (SmartNets) (pp. 1-4). IEEE. http://doi.org/10.1109/SmartNets58706.2023.10215825

[9] Livingston, L., Featherstone-Uwague, A., Barry, A., Barretto, K., Morey, T., Herrmannova, D., & Avula, V. (2025). Reproducible generative AI evaluation for healthcare: a clinician-in-the-loop approach. medRxiv, 2025-03. https://doi.org/10.1093/jamiaopen/ooaf054

[10] Patki, N., Wedge, R., & Veeramachaneni, K. (2023). The synthetic data paradox in clinical machine learning: Fairness distortions under minority class conditions. Artificial Intelligence in Medicine, 143, 102627. https://doi.org/10.1016/j.artmed.2023.102627

[11] Pilgram, L., Ko, H., Tung, A., & Emam, El. K. (2025). Protecting patient privacy in tabular synthetic health data: a regulatory perspective. NPJ Digital Medicine, 8(1), 732. https://doi.org/10.1038/s41746-025-02112-0

[12] Stadler, T., Oprisanu, B., & Troncoso, C. (2023). Membership inference in synthetic health datasets under realistic adversarial assumptions. Proceedings on Privacy Enhancing Technologies, 2023(4), 211–230. https://doi.org/10.56553/popets-2023-0064.